SIMULATION FOR DEDICATED LINE SMALL LOT SIZE MANUFACTURING

Wen-Yu Huang
Leo Ke
Tina Shen

Manufacturing Technology Center
Taiwan Semiconductor Manufacturing Company

8, Li-Hsin Rd. 6, Hsinchu Science Park
Hsinchu, Taiwan 300-78
REPUBLIC OF CHINA

ABSTRACT

With the speedup of product innovation, product life cycles are becoming shorter and shorter. To fully support customers, a foundry fab has to speed up its wafer-output schedule to shorten the time to market in the growth stage. High priority setting and small lot strategy are common techniques for accelerating manufacturing cycle time. However, small lot strategy would accompany 1.2% of extra equipment investment. The improvement should be sufficient so that the extra investment is worthy. Therefore, the novel strategy, dedicated line for small lot size manufacturing, is proposed in this paper, which allocates exclusive resources for expediting lots to greatly improve the manufacturing cycle time. By distinguishing dedicated lines from regular lines, the proposed strategy improves 21.7% of cycle time for expediting lots. Plenty of simulation results can evidence that the proposed strategy significantly outperforms mixed run mode.

1 INTRODUCTION

With the development of information technology and wireless communication, electronics industry grows rapidly with each passing day. Companies keep launching new, innovative and high performance products to enhance their brand equity and competitiveness. Therefore, production life cycles become shorter and shorter. For a foundry fab, fast manufacturing cycle time is one of key successful factors to meet customer requirements and help them to grab market share. Especially in the growth stage of product life cycle, how to rapidly manufacture the first batch of products with sufficient quantity would be crucial for customers to derive considerable revenue. The most straightforward method is to assign higher priorities to designated lots. In addition to priority setting, small lot size manufacturing (SLM) can also accelerate manufacturing cycle time, which divides designated lots into smaller lots. The common lot size in SLM is 12 pieces.

However, it is difficult to guarantee that small lots can always preempt manufacturing resources (e.g. tools) in mix-run environment. In fact, SLM benefits not only small lots but also normal lots in terms of cycle time. In other words, normal lots would also be accelerated by SLM. Therefore, it is still possible that pipelines of devices are fully occupied by normal lots. This unintentional status would reduce the performance of SLM. In this paper, a novel strategy named dedicated line for small lot size manufacturing is proposed to further improve the cycle time of SLM and solve the problem caused by mix-run environment. In the proposed strategy, partial resources form a distinguishing line which is dedicated to expediting small lots. As far as small lots are concerned, the regular line can be the backup of dedicated line. Small lot can switch between the regular and dedicated line to accelerate the cycle time and

2640

minimize the tool productivity loss. On the contrary, normal lots would only lie in the regular line to prevent from interfering with small lot operation. The access rules of the dedicated and the regular line for lots with different lot size are summarized in Table 1.

Table 1: Backup rules of dedicated line small lot size manufacturing.

Line Lot Size	Dedicated line	Regular Line
Expediting Small lots Lot Size <= 12	O	V
Unhurried small lots Lot Size <= 12	V	O
Unhurried normal lots Lot Size > 12	X	O

O: Default line; V:Backup line; X:Inhibited

2 SIMULATION MODEL

Autosched AP is adopted to validate the performance of the proposed strategy. 5 scenarios are involved in our study to reveal the effects of key variables, which are baseline, mixed run, dedicated line, utilization -5%, and utilization -10%. The baseline model is the typical case in a foundry fab. Both the mixed run and the dedicated line models adopt SLM to accelerate cycle time. In the mixed run model, lots with various lot sizes lie in the same line (regular line) and compete with each other. The dedicated line model is the proposed model which small lots have a dedicated line to insure their preemption. The utilization -5% model and the utilization -10% model are the variations of the baseline model. The setting of these models are shown in Table 2.

2.1 Derive theoretical process time and setup time from BIG DATA

When more small lots are in a fab, more setup time will be induced due to the frequent lot exchange. Theoretical process time and setup time of each lot size should be collected to judge the correctness of cycle time and the equipment capacity.

Generally speaking, a process time stored in MES is the difference between the track-in time and the track-out time. As shown in Figure 1, it is aggregated with the lot waiting time at the load port, setup time, and process time. In order to extract the setup time and the process time from the total process time, we decompose the process time recorded in MES by the following steps:

a. Collect all MES process time during a month
b. For a recipe, choose top 3 lot sizes appearing most frequently
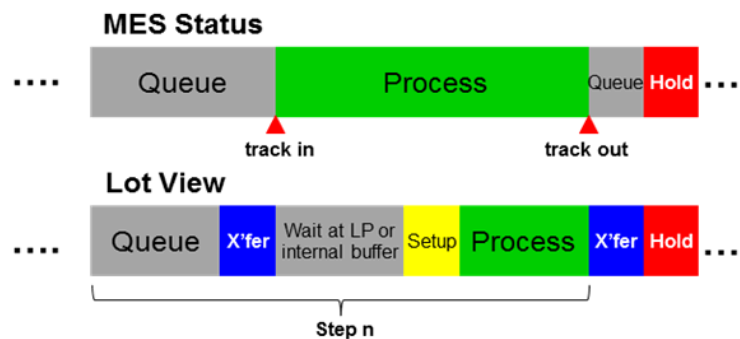

Figure 1: MES Process Time was aggregated with waiting at load port, setup time and real process time.

Table 2: Key independent and control variables of 5 scenarios.

	Variables	Baseline	Mixed Run	Dedicated line	Utilization -5%	Utilization -10%
Order	Qt'y - All	45000	45000	45000	42750	40500
	Q'ty – Normal line	45000	45000	30000	42750	40500
	Qt'y – Dedicated line	-	-	15000	-	-
	Part #	20	20	20	20	20
	Average Lot Size	1X	0.77X	0.77X	1X	1X
Equipment	Equipment Qt'y	1X	1.02X	1.02X	1X	1X
	NTB Qt'y	-	0.27X	0.17X	-	-
Operation Rules	Sampling Rate	1X	0.87X	0.87X	1X	1X
	Merge for Batching	N	N	30 mins	N	N
	Max Waiting Time Constraint	Y	Y	Y	Y	Y
	Hold Rate/Step	1X	0.9X	0.9X	1X	1X
	Hold Time	1X	1.1X	1.1X	1X	1X
	Transfer Time	1X	1X	1X	1X	1X

c. For each chosen lot size, calculate the slope of MES process time and percentile from 1% to 20% until the slope is less than 0.01 or reach 20%. The MES process time is near the theoretical process time of the lot size.
d. Adopt linear programming on the theoretical process time of 3 chosen lot size weighted by lot count, we can derive an equation.
e. The intercept on y-axis is setup time, and its slope is process time per wafer. The points above the equation are lot waiting time on load port.
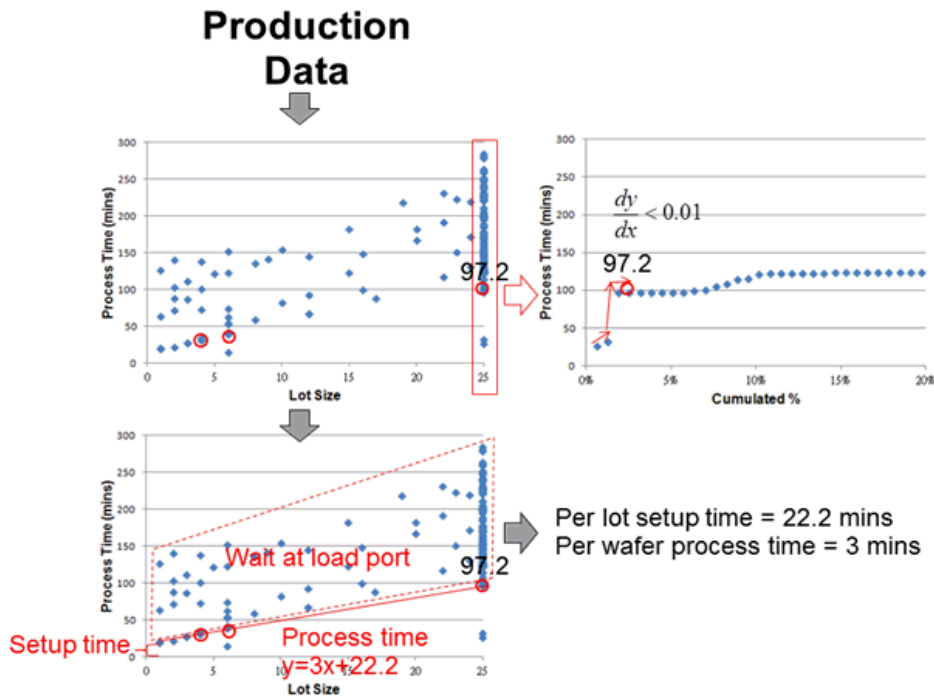

Figure 2: Example of extracting the setup time and the theoretical process time.

2.2 Model Architecture

As shown in Figure 3, the architecture of the simulation model is same as that of our existing model. However, the route and station files should be modified accordingly to react to more advanced devices adopting new technologies. Batch wafer start, priority, lot size distribution, and reshuffle are considered to meet the practical environment of a foundry fab. All these considerations can be realized in the order file. 20 key parts with different routes are selected to represent the actual product combinations. In order to obtain a precise process time, the number of load ports, pipeline time and setup group are taken into account in our simulation model. In addition, the distribution of the actual FOUP transportation time is modeled in the simulation based on the actual data in the fab.
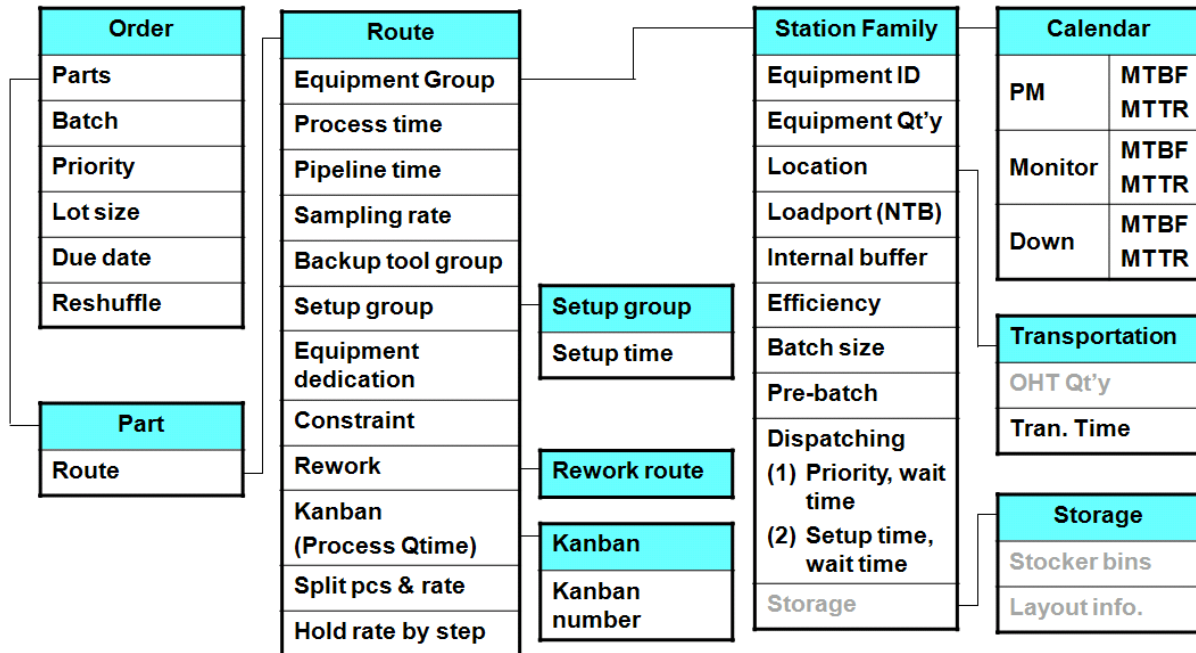
Figure 3: Model Architecture.

2.3 Baseline Model Validation

Baseline model is developed to faithfully reveal KPIs and actual status in an advanced foundry fab. The ramp up time of the model is around 210 days. After 210th day, a steady WIP (67250 pieces) can be obtained and the cycle time can be aligned to our target. KPIs are calculated based on the data from the 8th month to the 12th month. First, the WIP level of each module is compared to the actual WIP in a fab. Most of data are consistent except WIPs of module 8, 11, and 14. The reason is that there are more steps and longer process time in the stages of these modules. Second, we divide the whole route into 9 segments and investigate the WIP levels of each segment. The cumulated ratio of WIP is identical with our expectation. Third, the FOUP transportation time is also meet the actual KPI. According to the simulation results as shown in Figure 4, the baseline model is validated.

2.4 Mixed Run Model

In the Mixed run model, SLM is adopted to improve the cycle time. One-third of orders are in the form of small lot. The lot size of a small lot should be less than or equal to 12 pieces. By doing so, the average lot size is reduced by 23%. When the quantity of orders keeps the same, the total number of lots increases due to the adoption of SLM. In order to keep the total amount of wafers inspected by metrology tools, the

sampling rate should be reduced by 13%. Besides, the hold rate and hold time are increased by 10% in the mixed run model to behave that more errors occur with the increase in the total number of lots.

However, SLM accelerates the cycle time at the cost of productivity loss. This phenomenon is more serious for the tools with high throughput and long pipeline. To solve this problem, Near-Tool-Buffer (NTB) is introduced, which is a mini stocker adjacent to a tool. NTB is useful for shortening FOUP swap time and minimizing productivity loss due to the adoption of SLM. Even so, there is still 1.2% equipment investment needed to offset the productivity loss caused by SLM.
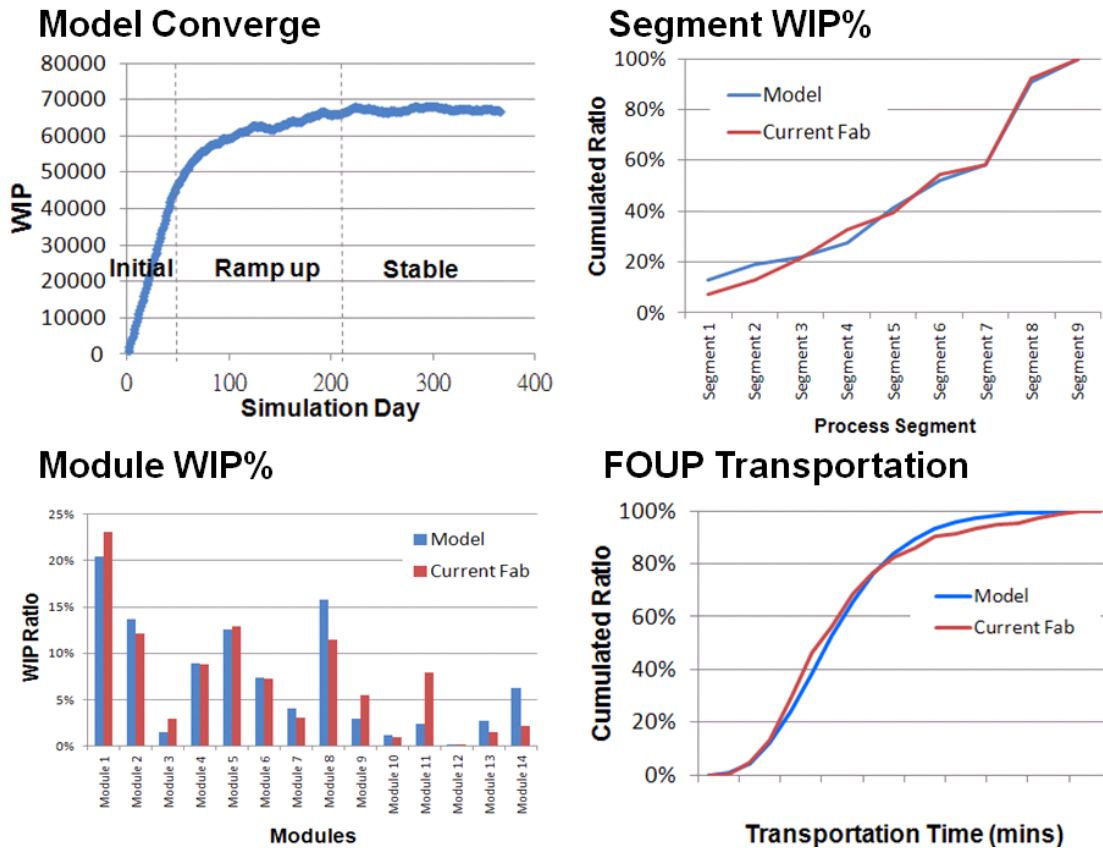
Figure 4: Baseline Model Validation.

2.5   Dedicated Line Model

The model of the dedicated line is similar to that of the mixed run model. The main different is that partial resources are designated and form a dedicated line for the manufacture of expediting lots (small lots). In addition, extra sorters are needed at the steps with larger batch size. In our model, 33% of tools are dedicated to expediting lots. The regular line is consisted of the other tools as usual. As shown in Table 2, 33% of wafers would be processed in the dedicated line. The process of these expediting small lots can be switched to the regular line if the waiting time exceeds 5 minutes. The small but not expediting lots can be switched to the dedicated line only if the waiting time exceeds 20 minutes. The normal lots (lot size >12) are prohibited from accessing the dedicated line. By doing so, small lots would not be interfered by normal lots. As mentioned previously, SLM would cause a certain productivity loss due to frequent FOUP exchange. Hence, 22 loops in the dedicated line have to equip extra sorters to recover the productivity loss. Small lots would be merged before these loops and split again behind these loops.

2.6 Utilization -5% and Utilization -10% Model

Obviously, the cycle time will be accelerated with the decrease of the utilization due to shorter queue delays. The utilization -5% model and the utilization -10% model are developed to demonstrate this phenomenon and examine whether SLM is cost-effective. The quantities of orders in the utilization -5% model and the utilization -10% model are 42750 and 40500 respectively.

3 RESULTS

Simulation results are summarized in Table 3. For mixed run model, the cycle time of expediting small lots has 13.1% improvement at the cost of 1.2% of extra tools. 7.2%, 4.3% and 1.6% are derived from run time, queue time, and hold time respectively. Normal lots also have 5.3% unintentional improvement due to the acceleration of small lots. The gap of the cycle time between expediting small lots and normal lots is 7.8%. The cycle time distribution of the mixed run model is shown in Figure 5.

The simulation results show the cycle time can be improved 21.7% by adopting the proposed dedicated line model. 10.9%, 7.7% and 3.1% are derived from run time, queue time, and hold time respectively. The same as the mixed run model, 1.2% of extra tools are needed to recover the productivity loss. The gap of the cycle time between expediting small lots and normal lots is 24.1%. Obviously, the proposed dedicated line model has a significant cycle time improvement for the expediting lots. The cycle time distribution of the dedicated line is shown in Figure 6.

The utilization -5% model and the utilization -10% model have 8.4% and 14% of improvement respectively in terms of cycle time. For the utilization -5% model, the 8.4% of improvement is consisted of 3.4% and 5% contributed by run time and queue time respectively. In the case of the utilization -10% model, the composition of the improvement are 5.7% and 8.3% derived from run time and queue time respectively. The results are sufficient to evidence that the dedicated line model is outstanding. The cycle time distributions of the utilization -5% model and the utilization -10% model are shown in Figure 7.

Table 3: Simulation Result Summary.

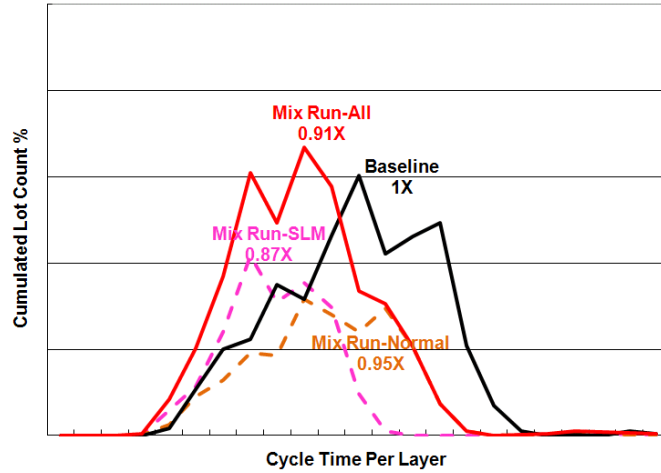| Simulation Cases	Average lot size	Cycle Time per Layer	Improvement on RQH				Extra Tool %
			Total	Run	Queue	Hold	
Baseline	1X	1X					
Mixed run model	1X	0.95X	5.3%	0.8%	2.8%	1.6%	
	0.53X	0.87X	13.1%	7.2%	4.3%	1.6%	+1.2%
Dedicated line model	1X	1.02X	-2.4%	-1.3%	-1.1%	0%	
	0.53X	0.78X	21.7%	10.9%	7.7%	3.1%	+1.2%
Utilization -5%	1X	0.92X	8.4%	3.4%	5.0%	0.1%	-
Utilization -10%	1X	0.86X	14.0%	5.7%	8.3%	0.0%	-

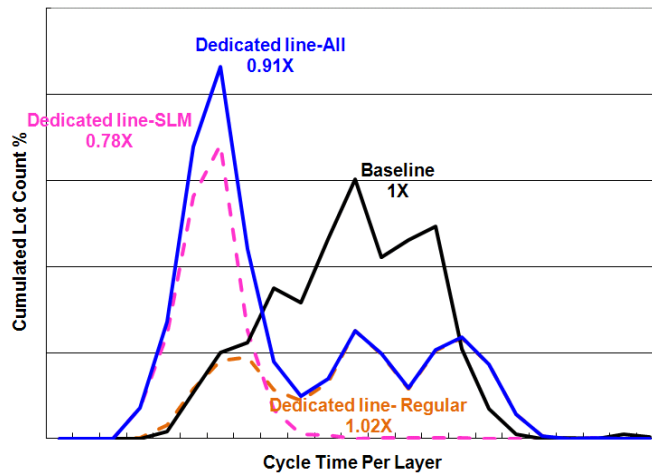Figure 5: Cycle Time Distribution of Mix-run.

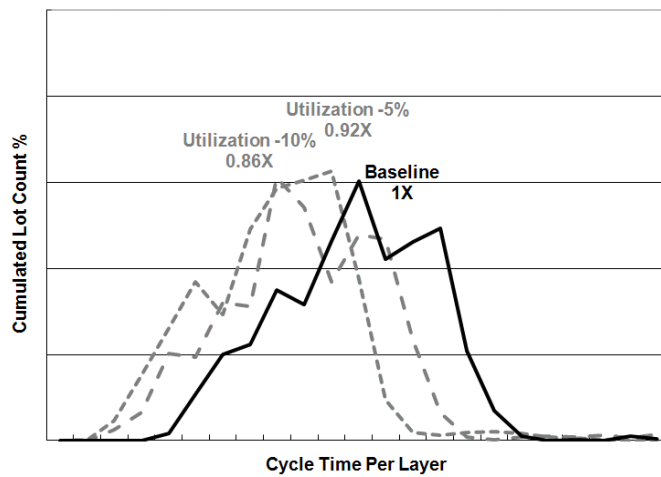Figure 6: Cycle Time Distribution of Dedicated line.

Figure 7: Cycle Time Distribution of Utilization -5% and Utilization -10%.

4   CONCLUSION

Aiming at the importance of the cycle time in the growth stage of products, accelerating the cycle time at the reasonable cost of productivity loss is one of major challenges for a foundry fab. This study evaluates the methodologies, benefit and cost to accelerate cycle time. Three strategies are evaluated in this paper: small lots in a mixed run environment, dedicated line for small lots, and low utilization environment. According to our investigations, only the strategy of dedicated line for small lots gains a considerable improvement in terms of the cycle time of expediting lots and minimizes the impact on productivity loss. In the proposed strategy, partial resources consist of a distinguishing line which is dedicated to expediting small lots. Expediting lots can utilize not only the dedicated line but also the regular line to accelerate the cycle time. The flexibility makes the cycle time faster and minimizes the impact on tool efficiency due to SLM. The simulation results can evidence the benefit of adopting the proposed strategy. The proposed dedicated line for small lots is useful to help customers to shorten the time to market and gain maximum profit. Some challenges are also identified in this paper such as flexible backup channels between dedicated line and regular line, APC rule refinement, and overhead reduction of lots, which are valuable information as applying SLM.

REFERENCES

Babbs, D., R. Gaskins. 2007. "Effectiveness of Small Batch Size on Cycle Time Reduction in a Conventional 300mm Factory." *Advanced Semiconductor Manufacturing Conference*, ASMC 2007. IEEE/SEMI, 105-110.

Kenyon, G., C. Canel, B. D. Neureuther, 2005. "The impact of lot-sizing on net profits and cycle times in the n-jobs, m-machine job shop with both discrete and batch processing." *International Journal of Production Economics*, 263-278.

Lee, H., C. H. Chien, H. H. Wu, Y. C. Chen. 2013. "Optimal lot sizing for 3DIC products in backend manufacturing," *2013 e-Manufacturing & Design Collaboration Symposium*, 1-4.

Miller, L., A. Bradley, A. Tish, T. Jin, J. A. Jimenez, R. Wright. 2011. "Simulating conveyor-based amhs layout configurations in small wafer lot manufacturing environments." *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 1939-1947, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Wang, C. N., C. H. Wang. 2007. "A simulated model for cycle time reduction by acquiring optimal lot size in semiconductor manufacturing." *The International Journal of Advanced Manufacturing Technology*, 34(9-10), 1008-1015.

Zhao, R., H. Ichimura, S. Takakuwa. 2013. "MFCA-based simulation analysis for production LOT-size determination in a multi-variety and small-batch production system," *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 1984-1995, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

Wen-Yu Huang is a technical manager in Manufacturing Technology Center, TSMC. He is in charge of fab simulation model design and development, equipment productivity analysis, and fab cycle time modeling. He received his B.S degree in applied mathematics and the M.S. degree in industrial engineering from National Chiao Tung University, HsinChu, Taiwan. His email address is wyhuange@tsmc.com.

Leo Ke is a technical manager in Manufacturing Technology Center, TSMC. He has 17 years experience in semiconductor industry. He received his B.S degree in computer science from National Chiao Tung University and MBA from National Taiwan University, Taiwan. His email address is lrke@tsmc.com.

Tina Shen is a deputy director in Manufacturing Technology Center, TSMC. She is TSMC assignee of ISMI (International SEMATECH Manufacturing Initiative) 450mm Program Advisory Group (450 PAG), Manufacturing Technology Program Advisory Group (MTP PAG) and ISSM, International Symposium on Semiconductor Manufacturing, International Program Committee. She won the Woman Innovated Award from the president of Taiwan in 2012. Her email address is syshen@tsmc.com.